DOCUMENT RESUME

ED 153 378 EC 110 037

AUTHOR Proger, Barton B.
TITLE Two Approaches to Common-Sense Program Evaluation in Special Education.
PUB DATE Mar 78
NOTE 14p.; Paper presented at the Annual Conference and Exhibition on Measurement and Evaluation (5th, Los Angeles, California, March 13, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Conceptual Schemes; *Evaluation Methods; Exceptional Child Education; *Handicapped Children; *Models; *Program Evaluation

ABSTRACT

Two models for program evaluation as related to decision making of routinely operated special education programs are discussed. One model deals with judgmental data collected from staff members once a year and ways to insert that data into the budget planning and program change processes. The other model deals with inclusion of student performance data on a program-wide basis and how to set expectancies for exceptional child performance. (Author/DB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED153378

TWO APPROACHES TO COMMON-SENSE PROGRAM EVALUATION IN SPECIAL EDUCATION

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Barton B. Proger

Coordinator of Program Evaluation Services
Montgomery County Intermediate Unit No. 23
Norristown, Pennsylvania

INTRODUCTION

It was with pleasure that I accepted an invitation to speak at this Fifth Annual Conference and Exhibition on Measurement and Evaluation. Although I began my educational career as a teacher, was trained later as a researcher and applied statistician, I have been functioning as a program evaluator for about ten years in connection with both routinely funded ("hard money") special education programs and federally funded ("soft money") special education projects. My purposes in this paper are to present two different models or approaches to routinely funded special education program evaluation: one dealing with student performance data and the other with administrative concerns. I shall also try to tie in many reactions to various versions of these two models as seen by supervisors, teachers, psychologists, and administrators.

It is also an interesting opportunity for me to examine how my own views as a program evaluator have been shaped by these ten years of practice. When I embarked upon this rather strange area of functioning, program evaluation as a discipline or profession was still embryonic, even in regular education practice. At that time we were being CIPP-ed or PERT-ed to death, and the model-builders and systems analysts were having a field day. My interest at that point was in applying some of the concepts in vogue to routinely operated special education programs (Proger, 1971a). Slightly later, my interests turned to more applied instructional decision-making in special education by means of criterion-referenced measurement (Proger & Mann, 1973). Throughout these ten years I have noticed a dearth of literature and other formal efforts in special education program evaluation. Also, two very basic problems appear to have plagued us in special education, with no easy resolution in sight yet: (1) how to make program evaluation data of any type truly relevant and not just a perfunctory academic exercise, and (2) how to set expectancy levels of performance not only for students (difficult enough)

EC110037

but also for program operations as a whole or any service component of them. It is my hope that the two models of special education program evaluation presented in this paper will address in some ways these latter two problems, and will stimulate some thinking among you. Before doing so, I will briefly give to you my perspective on the state of the art in program evaluation in both regular education and special education.

STATE OF THE ART

In the world of regular education or the world of program evaluation at large, we have seen several major developments. Division H for School Evaluation and Program Development of the American Educational Research Association has grown and prospered over the years. Several invitational conferences on program evaluation have been held in attempts to further the field. An example was the 1973 Conference on Evaluation sponsored by the University of Wisconsin at Madison (Apple, Subkoviak, & Lufler, 1974). Here one sees still further comments by such authorities as Scriven and Walberg. We now have gotten to the point where there are annual review volumes on evaluation (e.g., Glass, 1976) and, somewhat parallel to the Handbook of Research on Teaching, the Handbook of Evaluation Research (Guttentag & Struening, 1975). Needless to say, a number of articles have appeared in professional journals, and there is no point in reviewing them here, especially in view of the tremendous redundancy in them. (e.g., January 1978 issue of Educational Leadership).

Then we turn to special education. Here we have a wasteland. The only significant writings in this area have been in the health professions (e.g., Zusman & Wurster, 1975). Routinely operated special education programs in the public schools have not really been represented, other than occasional articles (e.g., Burrello, Tracy, & Schultz, 1973; Fairchild, 1975; Granger & Campbell, 1977; Guralnick, 1973; Haines, 1974; Proger, 1971b; Quay, 1973; Seymour & Proger, 1971; Vane, 1976; Vergason, 1973). Yet all of these writings are only peripheral with regard to the two crucial problems.

In general, when one examines both the general education positions on program evaluation and those of special education, he sees a lopsided emphasis on technical methodological techniques (e.g., research-oriented design considerations), the philosophical assumptions behind the discipline of program evaluation (e.g., whether or not a program evaluator passes judgments on the programs he is examining), instrument selection and construction, uses and abuses of instrumentation, and so on. When one sits back and examines these varied topics, he sees nothing more than the usual fodder of the formal researcher, with some qualifications dictated by the messy

world of routine program operation as opposed to controlled studies. These quasi-modified "textbook" considerations continue to be pushed in almost all writings in this area. "New" models continue to be produced, and yet in the world of dollars-and-cents decision-making, we still are plagued with two basic problems: how to make decisions based on program evaluation data, and how to set criteria for those decisions.

THE COMFORTABLE DILEMMA OF THE PROGRAM EVALUATOR AS A PROFESSIONAL

Program evaluators, on the one hand, and administrative decision makers, on the other, both head-nod to the fact that program evaluation data must be timely and relevant for practical decision-making. The most comfortable way a program evaluator can achieve this is to focus upon "one-shot" projects. The single biggest example is the world of federal projects. Knowing that, contrary to stated intent, most federal projects will be absorbed in only limited ways after federal funding ceases, and thus knowing in turn that there are no monumental financial decisions that have to be made by the sponsoring agency or LEA, the program evaluators and the key administrative personnel can settle back into the traditional niche of planning and applying good program evaluation design principles to the innovative project, collecting data at logically placed milestone dates, analyzing the data, and writing up and disseminating the project report. Since the consumers are implicitly understood to be the world-at-large (but not your own backyard!) eagerly awaiting the field-test results of the innovative program breakthrough that was supported by federal funds, there is no panic and pressure to link these results to a specific agency's decision-making timelines. In this setting, program evaluation is a very comfortable, very respectable operation. In fact, it can be exciting and very supportive of the program evaluator's self-concept. All of the tenets of program evaluation "fit" and "work," without undue pressure or criticism from those hard-headed administrative decision-makers. And, if your public agency is a large one, such as a city school district or regional educational service agency, this type of functioning is quite appropriate and indeed necessary, as dictated by federal reporting guidelines.

Now let's turn to the topic of my concern. While I too am quite comfortable when I deal with the above-described special purpose evaluations, I am ill at ease with how program evaluation data can be genuinely built into the decision-making processes of the locally funded, routinely operated programs. Needless to say, not much of this type of applied program evaluation occurs, let alone linking its results with day-to-day, pressure-ridden decision making. This is indeed the "seamier"

side of the failure of program evaluation, an oft-repeated theme. This is an uncomfortable area, one in which program evaluators (and for that matter, the key administrators who employ them) can be criticized for not taking full advantage of the possibilities of improving educational practice. For this "uncomfortable" area, I am proposing two different models: one for performance data (which I do not see being linked to administrative decision-making for many years to come, if ever at all), and one for judgmental data. Let's start with what I consider to be the more practical scheme.

MODEL A: JUDGMENTAL DATA

The basic assumption here is that administrative decision-making concerns override any desires for convenience or evaluative eloquence that the program evaluator might have. This means the information must be simple to use, constructive in nature, and be made available for preliminary budget planning. I am assuming here that if evaluative data is ever to affect routinely funded program practice, any such key decisions will of course be made largely on a financial basis, and most public agencies have internal budget hearing (or planning) activities that begin at some point during the midst of a fiscal year (in my specific case, this would be around February).

To allow major financial and program-wide decisions to be made (e.g., for all learning disabilities classes in an agency), the data must come to some decision-making group prior to the internal budget deliberations for the agency as a whole. In having an opportunity to observe many different data-based evaluation systems in action over these several years (day-to-day instructional decision-making data -- Proger & Mann, 1973, and Proger et al., 1972; program-wide, key-decision student data -- Proger, 1971b; accreditation-type, judgmental data -- Grotsky, Proger, 1975, and Grotsky et al, 1973), I am firmly convinced that the most appropriate and feasible data to provide to key decision makers for making major, program-wide changes, is judgmental data. It is futile to believe that various types of objective, performance data will ever "speak" in meaningful ways to decision makers or to evaluators; too much interpretation and inference is necessary, even with "clean" data. Since we are not comparing innovative program development with traditional practices (as in federal project or special purpose evaluation settings) but merely trying to carry out a routine, formative improvement upon existing, mass-produced special education, we need to rely upon the common-sense judgments. What better way to reflect in a small amount of information the experiential observations of the strengths and weaknesses of the program in day-to-day operation, than simple judgments from instructional and

administrative staff. The attached form serves this purpose. It is quite simple to use, to summarize, and to interpret. The form can be used at any time during the school year to allow linkage with the budget planning process. Every single staff member in the special education hierarchy would be surveyed for their opinions -- clinicians, psychologists, teachers, aides, and so on.

The form would be given out well in advance of the start of internal budget deliberations, but late enough into the school year to tap the problems experienced by various levels of staff. All forms from all programs would be returned to a single staff member (who might be a supervisor or other special education administrator), who might be designated a "Coordinator of Planning" for this purpose (no doubt an added duty to what they are already carrying out). The Coordinator would summarize the numerous responses within a program area (such as educable mentally retarded or speech). In my case, where budget planning commences in February, I would see the forms going out in mid-December and being compiled during January.

To guard against the data thus collected from becoming just one more item in a special education program supervisor's inbasket, a mechanism for building this data into the planning and decision-making processes must be built. What is suggested here is that the program supervisor be given the summary tabulation of responses by the Coordinator of Planning, with or without highlights in those responses as seen by the Coordinator. However, as appropriate, almost full responsibility for interpreting what changes, if any, are to be made in the given program will be made by that program supervisor (such as the supervisor of all educable retarded classes). The supervisor is to prepare a brief, written appraisal of his judgments of how these changes can be built into the budget process. The supervisor's written report would be discussed with the Director of Special Education. In this manner, the deliberations will be kept private but still will guarantee that the potential for change is being looked at, at the appropriate time in the year. (For example, during the next few years, the IEP process will probably foster a number of suggestions for change.) At the time of discussion with the Director of Special Education, the supervisor can bring all other sources of data into the discussion (such as more objective test data, parent input, or community input). The final decision-making is clearly left at the key levels where it should be. The data is only one means to an end, and the Coordinator of Planning is not directly involved in the decision-making process. Thus, logical decisions to change minor and/or major features of a routinely funded special education program can be reflected at a crucial time in the budget planning process. In turn, the Director of Special Education can reflect these desires for change (and their associated budget implications) with any of

his superiors.

Why judgmental data? Is it not too subjective and fraught with looseness? While these charges can (and always will) be made, the point is that they represent the distillation of several months of that staff member's feelings on the strengths and weaknesses of program operation as he views them every day under the pressure of the classroom or itinerant setting. To all of us who have at one time or another dealt with formal research and data collection of one kind or another, this all looks too simple and too commonplace. Yet, if you pause and ponder a moment, I think you will agree that here is real potential for program improvement. A current example may put this in perspective. Both regular and special educators across the country have been wrestling with how to implement Public Law 94-142 and the associated process of Individual Education Programs (IEPs). I have had the unique opportunity, with many mixed blessings, of implementing the IEP process in a county of 625,000 people. We have about 13,500 exceptional students in all of our special education programs (by comparison, Philadelphia has about 33,000 exceptional students). As most of you know, this process has had a profound influence on teacher contract negotiations, continuity of program, morale, Xerox costs, and our national forests! We are applying the data collection process I am describing here as "Model A" to this area to assist our special education administration to remold the IEP process and their day-to-day special education programs to each other so that 1978-1979 proceeds a little more smoothly. To show how crucial this simply collected data is, not only is it going to be used for internal budgeting and program-change purposes, but the Board of School Directors is also carefully examining the results.

One feature of "Model A" that has always appealed to me is its flexibility. There are a number of variations on how one links the data to the decision-making process. I mentioned earlier that the basic linkage is to have the Coordinator of Planning feed the data summary to both the program supervisor and the Director of Special Education, and to have them justify to each other why changes should (or should not) be made in various areas. Beyond this basic linkage, and prior to any such discussions of change, the program supervisors can really make their own instructional staff identify with the process if they create a small, manageable study team to review the data with the supervisor prior to his making any decisions to ask for programmatic and/or budgetary changes. I am sure you can devise other variations on this theme. Now on to the other model.

MODEL B: PERFORMANCE DATA

Earlier I posed the two basic issues of program evaluation: relevance (and impact) to decision making, and setting of

performance standards in a program evaluation sense. I think you can see how Model A affected the first problem, and that the second problem was addressed by mere discretionary standards from practical and financial viewpoints. However, the emphasis of Model A definitely was on the first problem. Now I shall give one scheme of trying to address the second concern primarily. I add as a footnote that it is somewhat ironic to me that I do not really see both basic problems of program evaluation being addressed equally well at one time within one model. There seems to be a delicate seesaw balance between meeting these two problems head-on. At any rate, the second problem of setting expectancy standards seems to be an appropriate area for the student performance data model, Model B, to tackle.

For several years, we have been dealing with how one interprets ongoing test data from exceptional students (see Proger, 1975, for an early attempt of this sort). I am thinking particularly of the case in which you are giving to all students in a special education program a common test (either a commercially produced normative measure, a commercially produced criterion-referenced measure, or some locally produced measure) at a common point (or set of time points) during the school year. The program may have several reasons for carrying out this type of uniform, mass measurement. First, there might be the intention of using the data in a program evaluative manner. Second, there might be the desire to have this data as just one of many types of input at an individual child level, to put that child's total battery of ongoing performance data in perspective (clearly, severe restrictions must be placed upon the mass-collected data when used in this fashion). Third, some claim that this type of data is helpful to communicate from special educator to regular educator when the child is ready to be integrated to any degree into regular education. There are many other splinter reasons as well. However, regardless of why one collects uniformly administered data on all special education students in a given program, the interpretation of that data for program improvement purposes poses a monumental problem. If one used a commercially produced normative instrument, the norms are meaningless for special education students. If one used a criterion-referenced device or some locally produced measure, there are no norms or standards that are useful in a program-wide, evaluative sense. The bottom line: how does one set expectancies? Before I address that issue, let me pose a decision-making mechanism for student performance data, just as we discussed for handling the judgmental data of Model A.

If any type of program-wide-collected and summarized performance data is to have any use at all in affecting program changes, there must be a formal mechanism for discussion of such data. We all know informal mechanisms (another name for nonaction) do not work. I propose that a standing formative evaluative committee be created within each program. The program supervisor would chair the committee. While I am concerned here

only with setting expectancies for student performance data, I could see this committee serving many other legitimate purposes. Among its measurement duties, the formative evaluative committee of each program would consider (a) the continuing appropriateness of the measures used in a uniform, program-wide manner; (b) what changes in program operation might be implied by changing patterns of such data; (c) how to set expectancies; (d) how often to measure. This committee would basically serve in an advisory capacity to the program supervisor. The committee would meet a few times during the school year, as often as needed. If available, a program evaluator or measurement specialist (such as a psychologist) would be appointed.

Now for the issue of expectancies. If you have large special education programs, then no doubt you are already collecting various types of administrative and student information for data processing by computer. We have such a system in Montgomery County that has been operating for a few years. All individually collected test data (such as generated by psychological records) and all uniformly, mass-generated test data (such as generated by program evaluation purposes) are loaded onto the computer. In the case of our system, we collect data on student performance twice a year (early fall and late spring). Sometime during summer, the computer services division analyzes the data according to whatever expectancy schemes we desire as administrators interested in program evaluation purposes.

In years past, we used to summarize the data in only descriptive fashion (grouping children within a program by age and/or intelligence and/or entering skill levels, and then generating corresponding statistics such as posttest levels of performance on various subtests, gains, standard deviations, and so on). I am of the opinion that all such data should be collected for descriptive purposes and that its potential for decision-making should be considered only very warily by the formative evaluation committee, as well as by any other groups or individuals. More recently, we have attempted to address the expectancy problem by using a variant on the theme of local norms. This scheme was an offshoot of a model posed by Ekwall (1972). Basically, Ekwall suggested that one look at changes in the rate of change itself. Ekwall had suggested that one compare current learning rate with the learning rate prior to entering a program. Ekwall is assuming one has pre and post measurements in terms of grade equivalents. One looks at how many years a child has been in school versus pretest grade equivalent levels, and computes a group average of prior learning rate. One then computes the group gain rate (sum of post scores minus sum of prescores) per student. The current gain rate is derived by dividing the group gain rate by the amount of a school year the children were in the special program under examination. Finally, the current learning rate is divided by the prior learning rate to get a measure of relatively how much faster the students are now learning in a special program.

In Montgomery County, we have now modified our existing computer listing and analysis programs to reflect Ekwall's procedure, as one more descriptive way to look at performance data. However, instead of computing group or class rates of learning, we have taken the approach of computing all learning rates on an individual student basis and then generating whatever cluster averages we wish from such individual data. By this individual child approach, one can (within the limitations of the flaws in the instrumentation) see how the child's current year of performance compares to the previous years. That comparison can deal with just the previous year, or all previous years, or any other logical way of looking at prior gain rate.

SUMMARY

Two models for program evaluation of routinely operated special education programs have been discussed. In both cases, a decision-making mechanism was described. One model dealt with judgmental data collected from staff members once a year and how to insert that data into the budget planning and program change processes. The other model dealt with how to consider student performance data on a program-wide basis and how to set expectancies for exceptional child performance.

Apple, M. W., Subkoviak, M. J., & Lufler, H. S., Jr. (Eds.) Educational Evaluation: Analysis and responsibility. Berkeley, Cal.: McCutchan, 1974.

Burrello, L. C., Tracy, M. L., & Schultz, E. W. Special education as experimental education: A new conceptualization. Exceptional Children, 1973, 40, 29-34.

Ekwall, E. E. Measuring gains in remedial reading. The Reading Teacher, 1972, Nov., 138-141.

Fairchild, T. N. Accountability: Practical suggestions for school psychologists. Journal of School Psychology, 1975, 13, 149-159.

Glass, G. V. (Ed.) Evaluation studies review annual. (Vol. 1.) Beverly Hills, Cal.: Sage Publications, 1976.

Granger, R. C., & Campbell, P. B. The school psychologist as program evaluator. Journal of School Psychology, 1977, 15, 174-183.

Grotsky, J. N., & Proger, B. B. Third-party evaluation of programs run from the state level: Two field-tested models. Educational Technology, 1975, 15, 44-49.

Grotsky, J. N., Proger, B. B., Hayden, D. L., Warkomski, R. C., & Murphy, W. Guidelines for Peer Evaluation Program (PEP): A systematic approach for evaluating educational programs. Harrisburg, Pa.: Pa. Dept. of Ed., Bureau of Special and Compensatory Education, August, 1973 (Available from ERIC as Document No. ED-083-304).

Guralnick, M. J., A research-service model for support of handicapped children. Exceptional Children, 1973, 39, 277-282.

Guttentag, E. L., & Struening, M. (Eds.) Handbook of evaluation research. (2 vols.) Beverly Hills, Cal.: Sage Publications, 1975.

Humes, C. W., II. School psychologist accountability via PPBS. Journal of School Psychology, 1974, 12, 40-45.

Proger, B. B. A formal program evaluation model for the special education programs in Pennsylvania. Harrisburg, Pa.: Bureau of Special and Compensatory Education, Pa. Dept. of Ed., 1971a. (available from ERIC as Document No. ED-108-396).

Proger, B. B. Program evaluation: The model building game. Journal of Learning Disabilities, 1971b, 6, 292-306.

Proger, B. B. Trends and patterns in achievement test results and baseline expectancy data for future performance: Programs for emotionally disturbed children, physically handicapped children, learning disabled children, and children with minimal brain injury. Annual program evaluation report, 1973-1974.

Proger, B. B., & Mann, L. Criterion-referenced measurement: The world of gray versus black and white. Journal of Learning Disabilities, 1973, 6, 72-84.

Proger, B. B., Mann, L., Burger, R. M., & Cross, L. H. Adapting criterion-referenced measurement to individualization of instruction for handicapped children: Some issues and a first attempt. Paper presented at 1972 Annual Convention of the American Educational Research

Association, Chicago, Ill., March, 1972 (Available from ERIC as Document No. ED-064-354).

Quay, H. C. Special education: Assumptions, techniques, and evaluative criteria. Exceptional Children, 1973, 40, 165-170.

Seymour, L. A., & Proger, B. B. Formal evaluation of the effectiveness of performance contracting programs for learning disabilities. In John I. Arena (Ed.) The child with learning disabilities: His right to learn. Selected papers on learning disabilities; 8th Annual International Conference, Pittsburgh, Pa.: Association for Children with Learning Disabilities, c. 1971. Pp. 75-84.

Vane, J. R. Problems in, and strategies for, evaluating preschool programs. Journal of School Psychology, 1976, 14, 39-55.

Vergason, G. A. Accountability in special education. Exceptional Children, 1973, 367-373.

Zusman, J., & Wurster, C. R. Program evaluation: Alcohol, drug abuse, and mental health services. Lexington, Mass.: Lexington Books (D. C. Heath), 1975.

ANNUAL SPECIAL EDUCATION PROGRAM EVALUATION

Directions: Each year we want to give each staff member an opportunity to express their concerns prior to budget preparation so that changes might be made wherever possible. Please continue on the back or on attached sheets of paper, if necessary. Thank you for your help.

1. What do you wish to offer in the following areas?

NATURE OF PROBLEM	RECOMMENDED SOLUTION
A.	
B.	
C.	
D.	
E.	

2. What aspects of program operation did you find the most satisfactory?

3. What other comments do you wish to make?